

Revisão 2: Sapatadas nas alturas

Felipe Figueiredo

March 30, 2016

Medidas sumárias

Alturas

Em episódios anteriores, fizemos uma coleta de dados na sala, dividindo-a em quatro setores α , β , γ e δ .

Setor α

O rol das alturas do setor α é:

162, 164, 165, 166, 167, 167, 171, 172, 173, 188

Podemos sumarizar os dados tanto como **média \pm desvio-padrão**, ou como **mediana \pm AIQ**.

Questionário 1

1. Esses dados são qualitativos ou quantitativos?
2. Discretos ou contínuos?
3. População ou amostra?
4. Determine a **Amplitude Total** dos dados
5. Sumarize os dados como **média \pm desvio-padrão**
6. Sumarize os dados como **mediana \pm AIQ**

Para as alturas deste setor, temos:

Média = 169.5

Desvio Padrão = 7.412452

Sumário 1: Média \pm DP = 169.5 \pm 7.4

Mediana = 167

Q1 = 165

Q3 = 172

AIQ = 7

Sumário 2: mediana \pm AIQ = 167 \pm 7

Setor β

O rol das alturas do setor β é

155, 160, 162, 168, 169, 171, 177, 179, 188, 190

Questionário 2

1. Determine a **Amplitude Total** da amostra
2. Sumarize a amostra como **média \pm desvio-padrão**

3. Sumarize a amostra como **mediana \pm AIQ**

Média = 171.9

Desvio Padrão = 11.6089046

Sumário 1: Média \pm DP = 171.9 \pm 11.6

Mediana = 170

Q1 = 162

Q3 = 179

AIQ = 17

Sumário 2: mediana \pm AIQ = 170 \pm 17

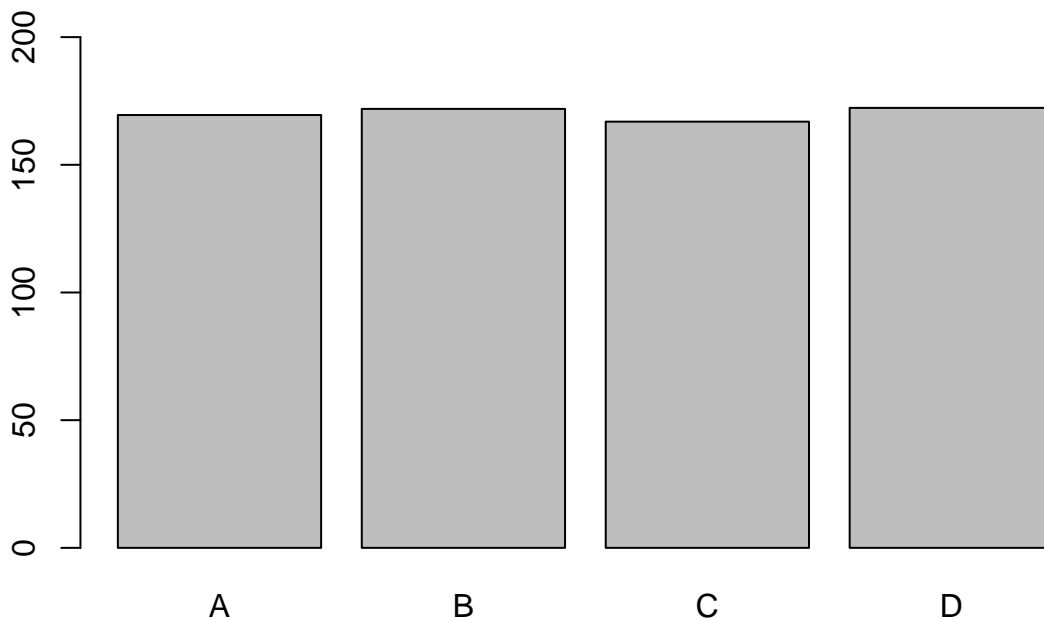
Comparando os setores α e β

Questionário 3

1. É possível concluir que as pessoas do setor α são mais altas que as do setor β considerando a média?
2. E a mediana?
3. E se você comparar o desvio padrão destes setores? Qual parece ter mais variabilidade?
4. E a AIQ dos setores?

Gráfico de barras das médias

Veja o gráfico de barras da altura média de todos os 4 setores.



Curiosidades

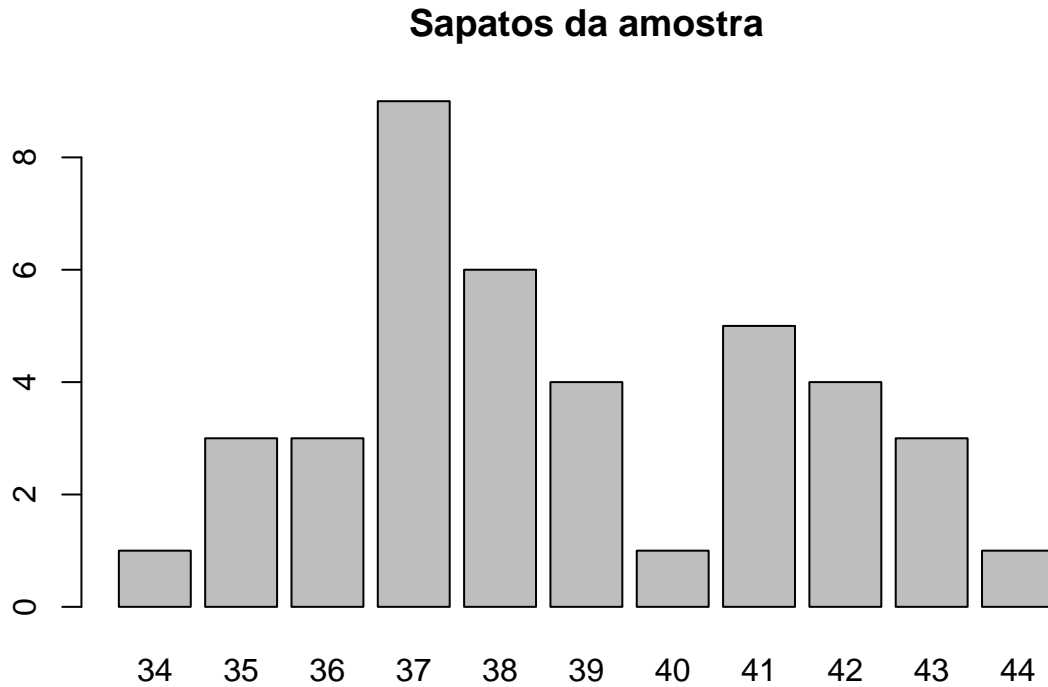
Como primeira curiosidade, vamos ver uma variação do gráfico de barras, quando queremos decompor a variável em subcategorias.

Sapatos

Vamos visualizar os tamanhos de sapatos, separados por sexo.

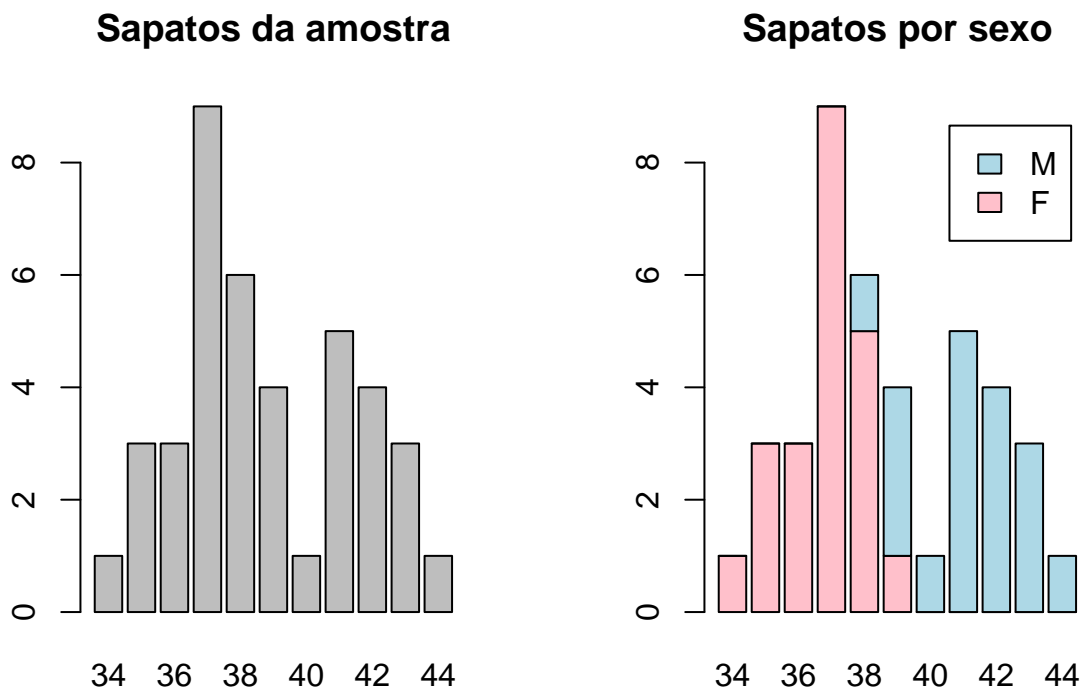
Primeiramente, o gráfico de barras dos sapatos, sem nenhuma subcategoria.

Gráfico de barras dos sapatos da amostra completa (40 pessoas)



Sem graça, né? Nem tanto! Dá pra ver que nesse gráfico tem dois “montes”, separados por um “vale”. Isso se chama **distribuição bimodal**, pois essas 40 amostras incluem tanto homens quanto mulheres, e ambos os sexos têm características diferentes. Essa versão cinza do gráfico já dá uma ideia que faz sentido procurar algum tipo de subdivisão nos dados, pra procurar distribuições unimodais.

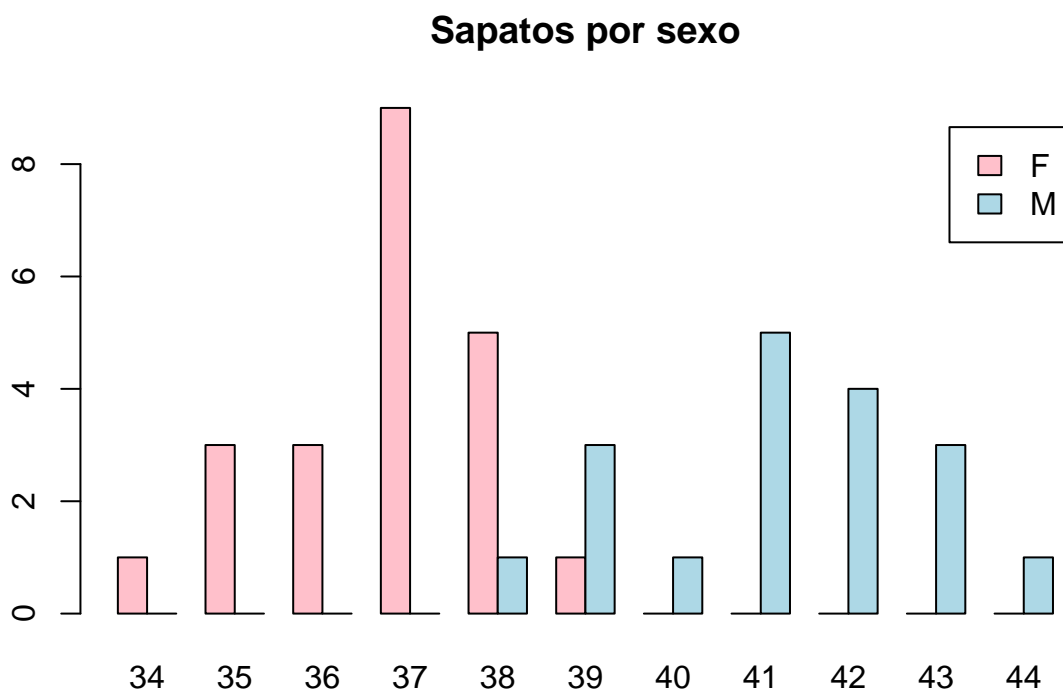
Gráfico de barras dos sapatos por sexo (empilhado)



Esse gráfico com os sexos empilhados não é utilizado com muita frequência, mas pelas cores já dá pra ver que as mulheres se estão nos tamanhos menores, e os homens nos tamanhos maiores (o que é muito razoável, sabendo que se trata de tamanhos de sapatos!).

Como é a maneira mais usual de representar esse tipo de gráfico de barras?

Gráfico de barras dos sapatos por sexo (separado)



Com barras independentes! Observe, no entanto, que agora as alturas das barras **não correspondem** às

alturas anteriores! Mesmo assim, fica bem mais fácil de saber quantas mulheres e quantos homens calçam cada tamanho, não?

Modelagem

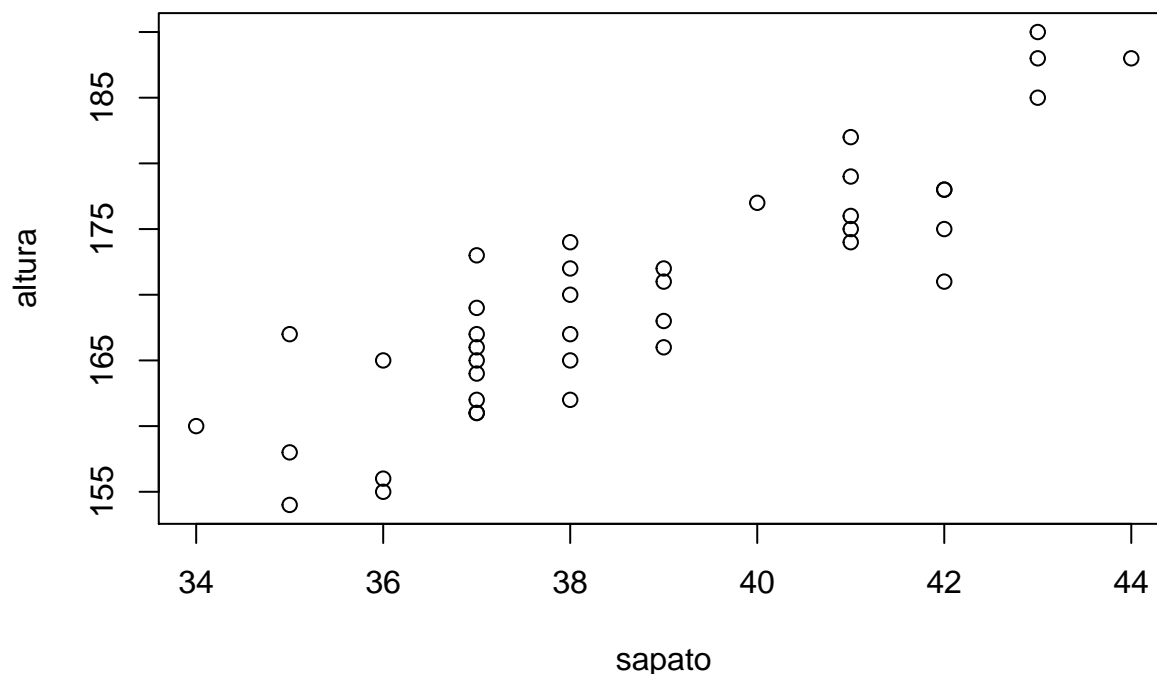
Como *curiosidade*, vamos ver como fazer um modelo, e a partir deste, fazer algumas previsões.

Sapatos x Alturas

Primeiramente, vamos visualizar os dados dados no plano cartesiano. Este tipo de gráfico é chamado de *gráfico de dispersão*.

Cada ponto representa uma pessoa (com seu sapato e altura, respectivamente).

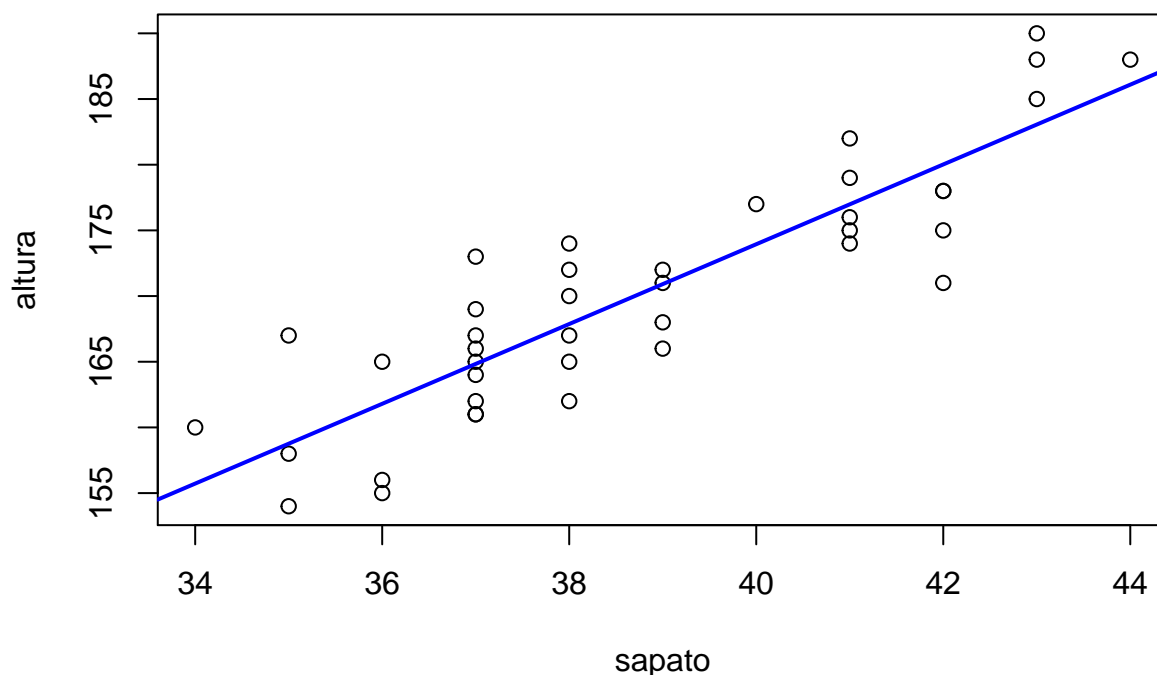
É possível observar uma *tendência* de crescimento, ou seja, quanto maior o tamanho do pé (eixo x), mais alta é a pessoa (eixo y).



Se quiséssemos *ajustar* uma curva aos pontos, que tipo de função seria apropriada?

Que tal uma reta?

Ajuste linear



Este **modelo** representa uma possível explicação para o fenômeno exibido nos dados da amostra. Se esta reta for uma boa explicação para os dados, podemos fazer previsões baseado nesta curva. Esta reta se chama *reta regressora*, já que usamos o método chamado Regressão Linear para encontrá-la.

Previsão numérica

Podemos por exemplo, estimar a altura de uma pessoa, sabendo quanto ela calça.

Por exemplo, se uma pessoa calça 42, a estimativa para sua altura é 180. Podemos visualizar esta *predição* no gráfico.

Na verdade, considerando a *margem de erro* desta estimativa, é provável que sua altura esteja entre os valores [177.8, 182.2]

Por outro lado, alguém que calça 36 deve medir 161.8, mas sua altura provavelmente se encontra no intervalo [159.8, 163.8].

Previsão categórica

Também podemos fazer estimativa de uma variável categórica, como por exemplo o sexo, baseado em quanto ela calça - ou mesmo se você vai passar, baseado na sua nota do primeiro bimestre (rsrs).

Se uma pessoa calça 42, a probabilidade de ela ser do sexo masculino é 99.997%. Agora se ela calça 36, a probabilidade já cai para 0.036%.